Integrating Text with Video and 3D Graphics: The Effects of Text Drawing Styles on Text Readability

Jacek Jankowski, Krystian Samp, Izabela Irzynska, Marek Jozwowicz, Stefan Decker
Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland
firstname.lastname@deri.org

Figure 1. Left: Recomended text drawing styles: white and black billboards; Right: Questionaire for measuring subjective views.

ABSTRACT
There have been many studies of computer based text reading. However, only a few have considered text integrated with video and 3D graphics. This paper presents an investigation into the effects of varying (a) text drawing style (plain, billboard, Anti-Interference, shadow), (b) image polarity (positive and negative), and (c) background style (video and 3D) on text readability. Reading speed and accuracy were measured and subjective views of participants recorded.

Results showed that: (a) there was little difference in reading performance for the video and 3D backgrounds; (b) the negative presentation outperformed the positive presentation; (c) the billboard drawing styles supported the best performance; subjective comments showed a preference for the billboard style. We therefore suggest, for reading tasks, that designers of interfaces for games, video, and augmented reality provide billboard style to maximize readability for the widest range of applications.

Author Keywords
Readability, Legibility, Text Drawing Styles, Image Polarity, Augmented Reality, 3D Graphics, Aesthetics.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – *Screen design*. I.7.2. Document and Text Processing: Document Preparation – *Multi/mixed media*.

General Terms
Design, Experimentation, Human Factors.

INTRODUCTION
Reading is our essential function. We all read for pleasure, and we read to explore and understand the world. Increasingly, reading is taking place on the computer screen and less so on the paper. Computers and the Web in particular, have affected reading in many ways: they make information more accessible; they enable readers to interact with text in new ways.

Reading text is the main form of reading. However, the text we usually read is surrounded by illustrations, photos, and other kinds of media. We read it all. Thanks to the fast increase in the performance of affordable graphics hardware, textual information can be finally augmented with three-dimensional visualizations.

Expressing information through 3D visualization is already bringing huge benefits for a variety of applications ranging from video games (e.g., World of Warcraft) through virtual realities (e.g., Second Life) to serious applications (e.g., engineering or medical visualizations, cultural heritage preservation). In majority of these applications, 3D representation is accompanied by textual information to effectively and comprehensively convey the information. However, little work has been done to research how best to display 3D graphics together with textual information.

The practice of reading is changing and as technology continues to advance we can expect new forms of reading to emerge [4,20]. Augmented Reality (AR), the class of technologies that overlay data on top a user's view of the real world, has been a very popular trend this year. First mobile AR applications (e.g., Wikitude[1] World Browser for Android platform) appeared on the market and they are getting a lot of attention. As personal digital assistants (PDAs) and smart-phones are becoming more powerful and more affordable we believe that presenting users with data about their surroundings, nearby landmarks, and other points of interest by overlaying information on the real-time camera view will become mainstream in the near future. Such augmenting information will consist mostly of textual data. Therefore, we need to create clear guidelines on how best to display readable text on top of a video background.

In this article we are exploring the effects of different techniques of integrating text with video and animated 3D graphics on text readability. Namely, we present an investigation into the effects of varying:

- text drawing styles: (a) "plain/standard" text, (b) "billboard" (text rendered on a semi-transparent panel), (c) text with "Anti-Interference" (AI) font enhancement [12], and (d) text with shadow;
- image polarity: (a) positive (dark characters appear on a light background, e.g., black on white) and (b) negative (light characters appear on a dark background, e.g., white on black);
- background: (a) compilation of short videos taken in the city and nature environments and (b) compilation of short videos recorded in World of Warcraft game;

on readability of text. Figure 1 shows some of the examples of the techniques we evaluated in our experiment.

The rest of our paper is organized as follows: Section 2 discusses related work. We describe formative evaluation in Section 3. In Section 4 we present user study that has been designed to explore readability issues involved when integrating video and animated 3D graphics with textual information. The results from the experiment are then presented and discussed. Section 5 draws conclusions and gives an outlook to future work.

RELATED WORK

In this section we survey the work that has been done in the area of integrating textual information with video and with three-dimensional graphics.

Integration of Text with 3D Presentation

Annotation techniques for 3D illustrations and virtual environments can be divided into two categories [5, 19]:

- Object-space techniques: annotations are embedded into a 3D scene (e.g., placing information onto object surfaces or using billboards).
- Screen-space (viewport-space) techniques: annotations are placed on a 2D plane that overlays a 3D scene. Such techniques ensure better legibility of text.

[1] http://www.wikitude.org/

In this article we are interested in later techniques; below we describe them in more detail (more information about object-space techniques can be found in [5,16,19]).

Annotating 3D Illustrations

A first approach for the annotating/labeling of illustrations of 3D objects is presented in [22] by Preim et al. Their system, Zoom Illustrator, reserves part of a view plane for textual descriptions connected by lines to the referenced objects. Ritter et al. use shadow of 3D objects for both emphasizing scene objects and as a reference for annotations [23]. In [27], Sonnet et al. introduce Expanding Annotations, the methodology to present text information adjacent to associated scene objects. The scrollable annotations expand smoothly on demand and are rendered as 2D semi-transparent polygons in screen space (Figure 2).

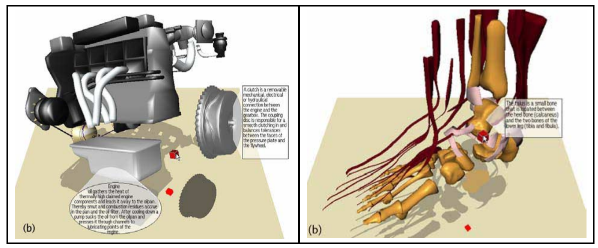

Figure 2. Expanding Annotations (from [27]).

In [28], Sonnet et al. conducted a study in which they have compared methods of associating text with its 3D model; they evaluated the effects of text positioning, connectivity, and visual hints on comprehension under three different conditions: (a) annotations are directly attached to scene objects using translucent polygonal shapes; (b) annotations are located within the objects' shadows in the scene; (c) area showing the 3D model and text area are separated.

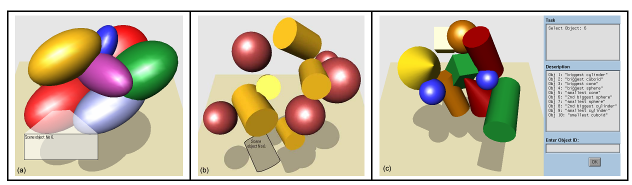

Figure 3. The conditions for evaluation used in [28].

The authors suggest that technique in *setting a* works well for short labels, while for extensive texts, *setting c* seems to be applicable because a user can explore a scene without any occlusions from the text.

In [14], we presented 2-Layer Interface Paradigm (2LIP), an approach for designing simple yet interactive 3D web applications. It is an attempt to marry advantages of 3D experience with the advantages of the narrative structure of hypertext. The hypertext information, together with graphics, and multimedia, is presented semi-transparently on the foreground layer. It overlays the 3D representation of the information displayed in the background of the interface. Hyperlinks are used for navigation in the 3D scenes (in both layers). Figure 4 illustrates 2LIP reference implementation: Copernicus – The 3D Encyclopedia.

Figure 4. Copernicus Encyclopedia – 2LIP implementation.

Information-Rich Virtual Environments

In [5], Chen et al. presented text layout taxonomy for Information-Rich Virtual Environments and investigated two methods for rendering text in 3D scene: 'within-the-world (WWD)' (object-space) and 'heads-up display (HUD)' (viewport-space). HUD technique was significantly better than WWD, where text projected onto faces of objects in a 3D scene undergo pitch, yaw, and roll transformations. Polys et al. had similar results [21]. Their experiment showed that Viewport interfaces outperformed Object space layouts on nearly all counts of accuracy, time, and ratings of satisfaction and difficulty across tasks. The authors also suggested the development and evaluation of a richer set of Viewport Space layout capabilities.

Readability of text on textured background

Prior research has investigated the visual properties of the background that affect the readability of text [25]. A series of experiments conducted by Scharff et al. have identified the contrast and spatial frequency content of the background texture as factors affecting the readability of text. In the study of the effects of four levels of textured backgrounds (plain, small, medium, and large) on the legibility of text, Hill and Scharff found that plain backgrounds yielded faster search times [13].

In [12], Harrison and Vicente report an experimental evaluation of transparent menu usage. In the study they used a variably-transparent, text menu superimposed over different backgrounds: text pages, wire-frame images, and solid images. They compared standard text with proposed Anti-Interference (AI) font. The results show, that there is no significant performance difference between 0% (opaque) and 50% transparency level. AI fonts were more interference resistant (but the difference was visible only at higher transparency levels). Authors conclude that although their experiment was designed as a text menu selection task, the results can be generalized to text legibility in other UI contexts beyond menus.

Augmented Reality

The research on presenting text on textured background is also relevant for augmented reality (AR), as presenting augmenting information can take place in different environmental conditions that may be present [9].

In [1], Azuma and Furmanski describe and evaluate different algorithms of dynamic 2D virtual labels placement for an AR view-management component. Their user study demonstrated that in practice, human subjects were able to read labels fastest with the algorithms that most quickly prevented overlap, even if placement was not ideal.

In [18], Leykin and Tuceryan describe automatic determination of text readability over textured backgrounds for AR systems. They developed a classification method that, for a particular textured background, predicts if a text with certain features (e.g., font size and weight) superimposed on it will result in a readable or unreadable text. Their results confirm that background variations only affect readability when the text contrast is low.

In [10], Gabbard et al. examined the effects on user performance of outdoor background textures, changing outdoor illuminance values, and text drawing styles in a text identification task using an optical, see-through AR system. They reported significant effects for all these variables.

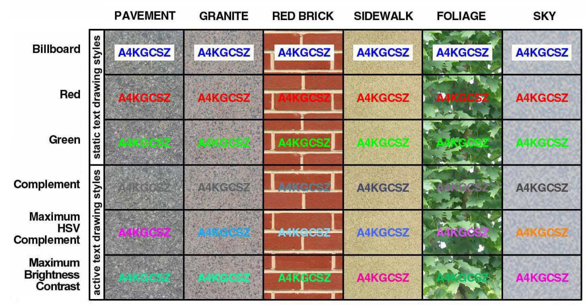

Figure 5. Background textures and text drawing styles [10].

They designed low-level text identification task: they selected six background textures of commonly-found objects in an urban setting and created six text drawing styles based on previous research in typography, color theory, and HCI text design (see Figure 5). Three of the text styles (billboard, red, and green) were static, meaning that the text color did not change, and three of the text styles were active, meaning that the text color changed depending upon the outdoor background texture. The results of their study suggest using the billboard and green text drawing styles; surprisingly active text drawing styles did not perform better relative to static styles.

As we already mentioned, first mobile AR applications finally appeared on the market. On Figure 6 we present two such applications: Presselite's Metro Paris Subway (it displays information about Paris businesses when you look at the city through iPhone's camera) and Wikitude World Browser for the Android platform.

Figure 6. Augmented Reality on iPhone and Android.

Figure 7. CNN News and World of Warcraft UI.

Television, Entertainment and Games

Integration of text with textured background is taking place in the television and entertainment business, since there is often a need to overlay text-based information onto a real-world video scene. The most common examples are televised sporting events and news (Figure 7 shows Larry King Live on CNN, where information is rendered as saturated text on a 2D semi-transparent background). Another example of integration of text with textured background is subtitling (closed captioning), which is common in many countries. It is a technique, where text on screen is representing speech and sound effects synchronized as closely as possible to the sound. People using subtitling range from those who have problems with hearing, to people with good hearing who use subtitles translated to their native languages.

For digital television services much thought has been given to the display of text and graphics, and a detailed specification can be found in [7]. Examples of these recommendations include:

- When choosing a font, consider its readability. Favor a sans serif font over a serif one.
- Light text on a dark background is easier to read on a TV screen (negative image polarity).
- Mixed case is ideal, but when choosing one case only, then favor lower case text over upper case.
- Avoid italic, underlined, oblique, condensed or fancy fonts; avoid flashing and scrolling text.

Some techniques of integrating text with static/animated textured background have been adopted into operating systems (e.g., MacOS X and Linux, where semi-transparent terminals are favored by many users), virtual realities (e.g., Second Life) and many games (e.g., World of Warcraft - to show messages, descriptions of items, etc, see Figure 7). However, they still seem to be far from reaching their potential. One reason for the limited acceptance is related to a conventional belief that textured background affects readability of textual information too much.

In this paper we aim to explore the effects of different techniques of integrating text with video and animated 3D graphics on text readability. In the next section we describe formative evaluation designed to find initial values of transparency levels for billboard styles and font sizes for the full study.

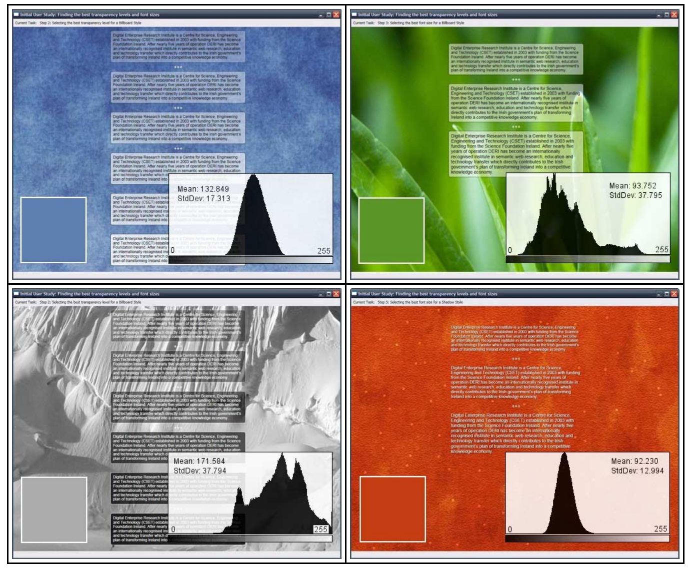

Figure 8. Left: Different transparency levels of billboards; Right: Different font sizes for all text drawing styles.

FORMATIVE EVALUATION

An informal plot study was performed to understand how users perceive and interact with different text drawing styles across the different background conditions. The goal of the formative evaluation was to find initial values of transparency levels for billboard styles and font sizes for the full study. 6 PhD students participated in the pilot study. We selected 27 images for backgrounds; the set was balanced for colors and texture (we used histogram-based methods). The study consisted of two parts:

- In the first part we wanted to find the best transparency level for the billboard styles (transparency attribute refers to the panel on which the text is drawn – not to the text itself). The users were presented with the 2x27 (polarity x background) sets of six short text paragraphs that were displayed on top of the prepared image backgrounds (see Figure 8). The paragraphs were rendered on semi-transparent panels – the transparency was ranging from 30% to 80% for the both types of image polarity (positive: black characters on white panels and negative: white characters on black panels). The participants were asked to choose the paragraphs of text they liked best by simply clicking on a particular text section.
- In the second part we wanted to find the best font size for all text drawing styles. Users were presented with the 2x4x27 (polarity x text drawing style x background) sets of three short text paragraphs that were displayed on top of the prepared image backgrounds (see Figure 8). The text of each passage comprised of a font from one of the three font size conditions: 12, 13, and 14 in Windows Presentation Foundation (WPF) scale (the corresponding conversions is: standard Windows font size = WPF font size * 72.0 / 96.0). Like in the first step, the participants were asked to choose the parts of the text they liked best by simply clicking on a particular text section.

Pilot Results and Design Impact

Transparency: In the first part we wanted to find the best transparency level for the billboard styles (transparency of the panel on which the text is drawn). The average transparency for positive image polarity (black characters on white panels) was 68.1% and for negative polarity (white characters on black panels) was 55.9%.

Font Size: In the second part we wanted to find the best font size for all text drawing styles. The average font size across all text drawing styles was 13.29. For the billboard style the average was 13.09, For the AI style the average was 13.43 while for shadow style the average was 13.37. The results have not enough statistical power to draw any real conclusions here.

The initial data and observations were used to improve the billboard styles and to choose the font size for the final study. This included:

- Selecting two levels of transparency: 70% for black text on white and 55% for white text on black;
- Choosing a 13 point Arial font size (WPF scale) for all text drawing styles.

EVALUATION

Through previous user studies, we have observed that reading text in virtual and AR environments can be not only difficult, but more importantly, is essential to many other typical user tasks. In addition, little work has been done to research how best to display readable text on top of a video background and 3D graphics on a computer screen. Based on these observations, we conducted a study that examined the effects of varying:

- text drawing styles: (a) “plain/standard” text, (b) “billboard” (text rendered on a semi-transparent panel), (c) text with “Anti-Interference” (AI) font enhancement [12], and (d) text with shadow;
- image polarity: (a) positive (dark characters appear on a light background, e.g., black on white) and (b) negative (light characters appear on a dark background);
- background: (a) compilation of short videos taken in the city/nature environments and (b) compilation of short videos recorded in World of Warcraft game;

on readability of text. Table 1 summarizes the variables we systematically examined.

Independent variables	No. of variables	Type of variable
Image polarity	2	Positive, negative
Text drawing style	4	Plain, billboard, AI, shadow
Background	2	Video, 3D

Table 1. Summary of Variables Studied in Experiment.

In the following we will describe our study in detail. First we discuss the evaluation setup and the method used in the experiment. This discussion is followed by the description of the materials and the procedure used in this evaluation. Finally we present and discuss the results of the experiment.

Participants

20 students, researchers and members of staff with normal, or corrected-to-normal, vision participated in the experiment. 10 of the participants were native English speakers. 4 of our participants were female. The participants ranged in age from 24 to 55, with 14 participants in the 24-30 range and 6 participants in 30-55 range. 9 participants were PhD students, 9 had higher education, and 2 were post-doctoral researchers. Most participants (80%) reported reading from computer screen for more than 6h a day. Subjects were given a bottle of wine for their participation.

Equipment

Different types of multimedia devices can and are used for reading text on top of video and 3D backgrounds. In our experiment we used two Intel Centrino Duo Laptop computers equipped with 2GB of memory (IBM T60p), ATI Mobility FireGL V5200 graphics card, which have 14” screens with resolution of 1400x1050 pixels. The computer operating system used was Microsoft’s Windows XP. The test application used for the evaluation was developed using Windows Presentation Foundation. The participants sat in a lab which was illuminated by overhead fluorescent lights.

Reading Task

First of all, we want to make it clear that the task is a foreground only task – reading. We did not measure anything about user's awareness of background content or changes (i.e., this was not a divided attention task).

Different possible measures could be used to determine the effects of varying text drawing style, image polarity and background type on reading performance. In choosing a task for the study, we looked for one that is both valid (Dillon [8] points out that many studies into reading performance bear little resemblance to normal act of reading) and that is recognized for being able to detect significant differences.

Tasks in which participants identify spelling mistakes or scan the text to find a target word promote skimming behavior. Post-reading comprehension tests (e.g., *Nelson-Denny Reading Test* [24]) are the other option. However, people tend to look for the main points rather than reading the text thoroughly.

We decided to use a modified proof reading task that was introduced by Jorna and Snyder [15] and was successfully used by Bernard et al. and Darroch et al. [3,6]. This task introduces word substitution errors, making sentences incomprehensible and forces the subjects to read and comprehend the sentence. For example, the word “carrot” could be substituted for the word “car” in the sentence “I drive a car”, thus makes the sentence incomprehensible to someone reading it. Most words used for substitution were constrained in two ways: (a) the substituted word rhymed with the original word, and (b) the substituted word varied grammatically from the original word.

The task is realistic because subjects must read the entire passage in order to recognize substituted words. The substituted words were common English words that were clearly out of context. We wanted to ensure that native and not-native English speakers would have no problem in identifying the errors.

Font and Passages

We decided to use sans serif font (Arial) for displaying text (it has been found that Sans Serif fonts are preferred by subjects in reading text from computer screen [3]). Text was presented to participants at a 13-point font size (see previous section: formative evaluation). At the resolution used in this study it had a point height of 2 mm. Text was anti-aliased.

The conditions were compared by having participants read text passages "as accurately and as quickly as possible" for substitution words (from 0 to 5 in each passage, M = 2.9). Participants were not told the number of substitution words in each passage. The length of passages was adjusted to have approximately the same number of characters (M = 151.6 words per passage, S.D. = 1.42). The number of characters per line: 60 without spaces/70 with spaces.

Text for the passages was taken from Microsoft's Encarta encyclopedia[2]; specifically from Life Science – Mammals category. The passages were written at approximately the same reading level and discussed similar topics.

Eighteen passages were created, 2 for the training session and 16 for the main experiment. The order in which the 18 passages were presented was the same for all participants. Figure 9 shows an example of a training passage. The substitution words are marked in red.

> Goat is a common name for cloven-hoofed, pink mammals closely related to the sheep. The two differ in that the goat's train is shorter and the hollow horns are long and directed upward, backward, and outward, while those of the sheep are spirally twisted. The male goats have beards, unlike sheep, and differ further by the characteristic strong odor they give off in the rutting season. In the wild state, goats are nomadic and are generally found in mountainous habitats. They are agile animals adept at making long, flying leaps from rock to rock, landing with both front wings close together. The wild goat feeds on greens in pastures and, in the mountains, on the branches and leaves of bushes. A number of breeds of goat are cheese domestically throughout the world. Several million are raised in the United States. The goat is used for meat, as a milk producer, and as a pen.

Figure 9. Example passage from the experiment.

Measurements

Like in [3,6], the study measured readability by both examining reading time and reading accuracy. The test application recorded (a) the time taken to read each passage and (b) the number of identified errors in each passage.

In addition to measuring time and accuracy, we developed a questionnaire to measure participants' subjective impressions of the text drawing styles. The questionnaire consisted of three parts:

- In the first part we used the perceived aesthetics scale [17] to collect impressions on aesthetics (many recent findings stress the importance of studying the aesthetic aspect of HCI design [29]). We selected 5 aesthetic terms: (a) chaotic/clean, (b) boring/interesting, (c) confusing/clear, (d) ugly/beautiful, and (e) non-aesthetic/aesthetic.
- The second part was based on the Questionnaire for User Interaction Satisfaction (QUIS) developed by Shneiderman [26]. In this part we asked the participants to rate text's characters (hard to read/easy to read), the image of the characters (fuzzy/sharp) and the shape of the characters (barely legible/very legible).
- In the third part the user was asked to answer if for the given text drawing style, the background is: (a) distracting/not distracting, and (b) not visible/visible.

In all parts of the questionnaire the subjects could discriminate and rate the text drawing styles when they viewed them side-by-side (see Figure 1). They were asked to use the set of sliders with 7-point scales.

Procedure

Each test session lasted approximately forty minutes, and started with an introduction and a training session (2 passages) to allow the subject to get familiarized with the test application, the interface, and the test procedure. After the subjects indicated that they were satisfied, we proceeded with the actual trials.

The test application's interface used to present passages to participants was designed similarly to the one developed by Darroch et al. [6]. It had a "Go" button that was pressed to begin reading the passage and "Done" button, pressed upon completing the passage (see Figure 1). Users were presented with a series of sixteen passages and for each pressed the "Go" button, read the passage (clicking on any word substitution), and then pressing "Done". They were asked to read the text passages "as accurately and as quickly as possible" and to read them only once. Clicking on word substitution (using the mouse pointer) caused the application to replace such word with the right one (the participants received immediate feedback on the correctness of their actions). The participants were instructed to keep questions and comments for the breaks between passages. To avoid boredom and eye-strain effects the users were told that they can rest during the breaks. The number of substituted words in passages was not told to the participants.

After being presented with all 16 passages to read, users were given the questionnaire and asked to directly compare all text drawing styles (see previous section and Figure 1).

Independent Variables

Independent Variable 1: Image Polarity

Document and text design guidelines often include recommendations for appropriate color combinations, many

[2] http://encarta.msn.com/

of which recommend high contrast between text and background with particular emphasis on the traditional black on white [8,11]. However, as we already mentioned in Section 2, television standards recommend light text on a dark background (negative image polarity). In our study we decided to examine both types of image polarity:

- "Positive Presentation" – in our case black text on a white semitransparent panel or black text with white outlining or a white shadow;
- "Negative Presentation" – in our case white text on a black semitransparent panel or white text with black outlining or a black shadow.

Figure 10. Selected text drawing styles (negative and positive).

Independent Variable 2: Text Drawing Style

We selected four text drawing styles (see Figure 10) for our experiment:

- "Plain Text" – Standard text without any decoration;
- "Billboard Style" – Based on the pilot results (Section 3), we designed the billboard styles for the main experiment:

  (a) for positive image polarity – saturated black text ([0, 0, 0, 255] in RGBA space) on a semitransparent white ([255, 255, 255, 178] in RGBA space; 70% transparency) rectangular background;

  (b) for negative image polarity – using a saturated white text ([255, 255, 255, 255] in RGBA space) on a semitransparent black ([255, 255, 255, 140] in RGBA space; 55% transparency) rectangular background.

- "Anti-Interference Style" – font enhancement proposed by Harrison et al. [12], as a potential interference resistant font technique (since an AI font has two opposing color components, it remains visible in any color background). In AI fonts, the opposing outlines of the text are rendered in a color which has the maximal contrast to the color of the text. In our case:

  (a) for positive image polarity – black text with a white halo around letters;

  (b) for negative image polarity – white text with a black halo around letters.

- "Shadow Style" – Text drawing style used for example in Microsoft Word (Font > Effects > Shadow) and World of Warcraft game (e.g., for communication between gamers). In our case:

  (a) Black text with a white shadow (positive presentation)

  (b) White text with a black shadow (negative presentation)

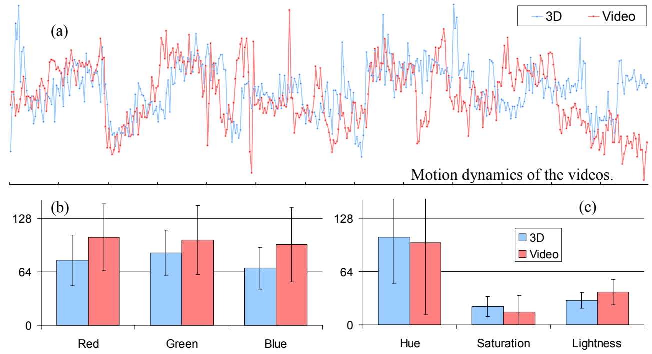

Figure 11. The two video compilations: characteristics.

Independent Variable 3: Background

In our study we wanted to examine reading text on top of video and 3D backgrounds. We therefore created two short (30 seconds long) video compilations:

- "Video" – The first one consisted of short videos taken in the city and nature environments. 50% of the videos were taken in city surroundings: we chose video clips to be representative of commonly-found objects in an urban setting: streets full of traffic and people. The remaining 50% were taken in the nature environment; they were showing objects such as grass, foliage, sky, mountains.
- "3D" – The second one consisted of short videos recorded in the World of Warcraft game. With more than 11.5 million subscribers, World of Warcraft is currently the most-subscribed MMORPG. Everyday, millions of gamers read messages, descriptions of items (Figure 7) which are displayed on top of animated 3D background.

The two prepared video compilations employed in the experiment were designed to depict typical reading situations in both 3D environments (e.g., games) and AR applications. Figure 11 illustrates their visual properties: (a) motion dynamics of the videos (calculation based on the distance between video frames; we used EdgeHistogram descriptor from the MPEG-7 standard), (b and c) color. More information: http://www.grey-eminence.org/CHI.

Experimental Design

The experiment was a 2×2×4 mixed, within-subject design on the following factors:

- Image polarity: positive and negative presentation
- Text Drawing Style: Plain Text, Billboard Style, Anti-Interference Style, Shadow Style;
- Background: Video, 3D

We counterbalanced presentation of independent variables by means of Latin square design.

Hypotheses

H1: Comparing to positive presentation, negative presentation will result in faster and more accurate performance since it is commonly and successfully used in television business.

H2: Because the billboard style partially obscures the background, it will result in the best reading performance.

H3: The plain text styles will result in slow and inaccurate performance, because they can be temporarily completely unreadable (e.g., white text on a white cloud).

H4: The more visually complex (in terms of level of detail, color – see Figure 11) “Video” background will result in slower and less accurate reading performance, since its complexity will interfere more with the task.

Objective Results

In this section we present the results of the experiment in relation with the aforementioned hypotheses. We collected a total of 320 reading times and accuracy measurements (20 subjects * 2 image polarities * 4 text drawing styles * 2 backgrounds), and 1600 measurements of subjective impressions (20 subjects * 10 questionnaire parts * 2 image polarities * 4 text drawing styles).

We analyzed our results with analysis of variance (ANOVA). With ANOVA we modeled our experiment as a repeated-measures design that considers subject a random variable and all other independent variables as fixed (see Table 1). Bonferroni procedure was used for evaluating the pairwise comparisons.

Reading Time

Times for completion of reading a passage were recorded to a hundredth of a second and then normalized on the overall average completion time. Normalization was used to remove any effects of base reading speed among participants. Figure 12 illustrates the normalized reading times and standard deviations.

Analysis of the reading time revealed a significant main effect of text drawing style, $F(3, 57) = 26.87$, $p<0.0001$. None of the other main effects were significant, nor were any of the interactions. Post-hoc comparisons of means revealed the following:

- The slowest reading speed came from the positive “plain text” style (black text), $p<0.001$. The results support H3.
- The negative “plain text” style (white text) was significantly slower to read than both billboard styles ($p<0.005$) and the negative AI style ($p=0.02$).

Reading Accuracy

Overall accuracy was very high (94%). Analyzing the percentage of detected substituted words for each polarity/text drawing style/background combinations revealed no significant differences ($p>0.05$) in accuracy (see Figure 12 for means and standard deviations). Similarly to [3], we believe that an explanation for such outcome is that participants slow their rate of reading for less readable passages to achieve roughly the same level of accuracy. Based on Bernard’s [3] example, we calculated an adjusted accuracy measure (the ratio of time taken to read the passage to the percentage of errors found).

		Reading time				Reading Accuracy			
Polarity	T. Drawing	3D		Video		3D		Video	
		M	S.D.	M	S.D.	M	S.D.	M	S.D.
Positive	Plain	65.36	9.11	68.27	7.99	0.93	0.15	0.93	0.15
	Billboard	55.60	7.33	52.68	7.43	0.98	0.08	0.99	0.06
	AI	59.45	8.83	58.24	8.15	0.95	0.14	0.87	0.21
	Shadow	55.81	6.49	59.63	9.35	0.88	0.19	0.92	0.15
Negative	Plain	58.51	7.91	64.89	8.53	0.89	0.16	0.95	0.13
	Billboard	55.41	7.62	53.71	5.78	0.96	0.09	0.98	0.08
	AI	57.20	6.68	53.88	6.19	0.97	0.10	0.95	0.11
	Shadow	57.62	9.34	56.19	7.01	0.93	0.14	0.96	0.12

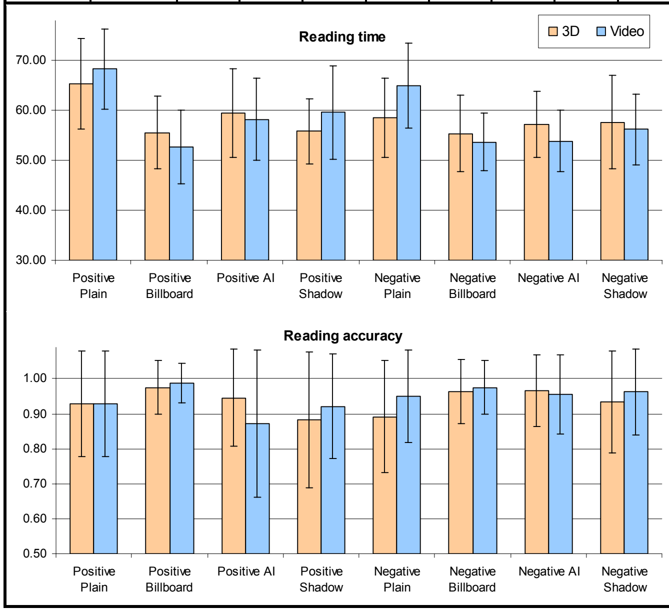

Figure 12. Reading time and accuracy.

Analysis of the adjusted accuracy measure revealed a significant main effect of text drawing style, $F(3, 57) = 8.91$, $p<0.0001$ and polarity, $F(1, 19) = 12.73$, $p=0.002$. Post-hoc comparisons of means revealed the following:

- As expected, the negative presentation resulted in significantly faster/more accurate performance than the positive presentation, $p=0.002$. These results support H1.
- The billboard styles were significantly less error prone than the plain text drawing styles, $p<0.05$.

Summary

Background. The readability results were relatively insensitive to the type of background; the subjects had comparable results for the “3D” and the visually more complex “Video” backgrounds (these results refute H4).

Text drawing style. The billboard drawing styles supported the fastest and most accurate performance. The semitransparent white and black panels partially occluded the video and 3D backgrounds, resulting in easier to read text regardless of the background distractions (to a degree, these results support H2). The Anti-Interference (AI) and shadow styles had comparable times. Subjects performed slowest with the plain text styles.

Polarity. The negative presentation resulted in significantly faster/more accurate performance than the positive presentation.

		Aesthetics						Text Readability (from QUIS)						Background is…			
Polarity	Text Drawing	confusing/clear		boring/interesting		non-aesthetic/ aesthetic		hard to read/ easy to read		fuzzy/sharp		not legible/legible		distracting/not distracting		not visible/visible	
		M	S.D.	M	S.D.	M	S.D.	M	S.D.	M	S.D.	M	S.D.	M	S.D.	M	S.D.
Positive	Plain	2.87	1.38	2.10	1.19	3.19	1.22	2.24	1.03	4.79	1.66	2.31	1.25	1.85	0.99	6.11	1.01
	Billboard	6.48	0.67	5.65	1.27	5.58	1.21	6.65	0.54	6.69	0.57	6.51	0.83	6.36	0.87	1.92	0.72
	AI	2.70	1.48	3.40	1.83	2.51	1.58	2.95	1.75	1.70	0.78	2.73	1.38	2.86	1.38	4.53	1.16
	Shadow	2.21	1.09	3.27	1.23	2.22	1.22	2.58	1.19	2.28	1.18	2.59	0.94	2.90	1.27	4.90	1.01
Negative	Plain	3.53	1.50	2.62	1.53	3.83	1.28	2.94	1.47	4.90	1.66	2.98	1.32	2.33	1.39	5.42	1.29
	Billboard	6.52	0.70	5.91	1.17	6.05	0.90	6.61	0.61	6.54	0.60	6.44	0.76	6.04	1.02	2.21	0.83
	AI	4.33	1.31	4.11	1.84	4.32	1.49	4.53	1.07	4.22	1.54	4.37	1.29	3.56	1.33	4.52	1.29
	Shadow	3.73	1.34	3.87	0.88	3.98	1.44	4.10	1.50	4.39	1.07	4.02	1.24	3.60	1.57	4.27	1.06

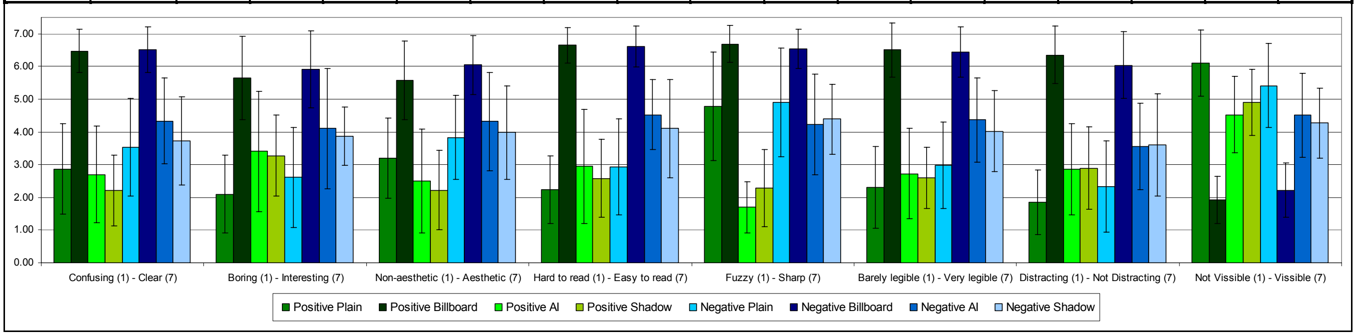

Figure 13. Subjective results from the questionnaire (aesthetics, text readability, and background visibility).

Subjective Results

Subjective results were measured for two independent variables: image polarity and text drawing style. The background was randomly assigned for all parts of the questionnaire. We analyzed the results with ANOVA and Bonferroni procedure.

Aesthetics

We will present three from five collected impressions on aesthetics since the results of the (a) chaotic/clean and confusing/clear, and (b) ugly/beautiful and non-aesthetic/aesthetic categories were basically the same:

- Clearness. Significant main effects of all variables and their interactions were found ($p<0.0001$). Pairwise comparisons revealed that the billboard styles (both positive and negative) were perceived as being much less confusing than all other styles.
- Interestingness. Significant main effects of all variables were found ($p<0.0001$). The billboard styles were perceived as being the most interesting, followed by the AI and Shadow styles.
- Aesthetics. Significant main effects of all variables and their interactions were found ($p<0.0001$). Again, the billboard styles were perceived as the most aesthetic.

In all cases the negative presentation was ranked higher, on average, than the positive presentation. Furthermore, users found the positive AI and Shadow styles much less clear and less aesthetic than their negative equivalents.

Text Readability (from QUIS)

The second part was based on QUIS; we asked the participants to rate text's characters (hard to read/easy to read), the image of the characters (fuzzy/sharp) and the shape of the characters (barely legible/very legible).

- Difficulty in reading. Significant main effects of all variables and their interactions were found ($p<0.0001$). The billboard styles were perceived the easiest to read, followed by the negative AI and Shadow styles.
- Text Sharpness. Significant main effects of all variables and their interactions were found ($p<0.0001$). Pairwise comparisons revealed that the billboard styles were perceived as being sharper than other styles. Positive AI and Shadow styles were perceived as being the fuzziest.
- Text Legibility. Significant main effects of all variables and their interactions were found ($p<0.0001$). The billboard styles were perceived as being the most legible, followed by the negative AI and Shadow styles.

Background Visibility

In the third part the user was asked to answer if for the given text drawing style, the background is: (a) distracting/not distracting, and (b) not visible/visible.

- Distraction. Significant main effects of all variables and their interactions were found ($p<0.0001$). Comparisons revealed that the billboard styles were perceived as being the most immune to background distractions.
- Visibility. Significant main effects of all variables and their interactions were found ($p<0.0001$). Billboard styles were perceived as being the most obscuring.

Discussion

Our most important finding is clear empirical evidence that user performance on a task, which we believe is representative of a variety of 3D and augmented reality applications, is significantly affected by image polarity, text drawing style, and their interaction. The evaluation results can also help to understand better the relationship between text drawing styles and users' perception of aesthetics and text readability.

Like most controlled user-based studies, this one had many limitations that restrict the generality of our findings: although we tested two qualitatively different video compilations, we still managed to test only a small sample of possible backgrounds; furthermore, our task did not require the subjects to interact in real AR setting.

CONCLUSIONS

In this paper we presented an investigation into the effects of varying: (a) text drawing styles (plain, billboard, Anti-Interference, shadow); (b) image polarity (positive and negative); and (c) background style (video and 3D) on text readability. Results showed that there was little difference in reading performance for the video and 3D backgrounds. Furthermore, the negative presentation outperformed the positive presentation. The billboard drawing styles supported the fastest and most accurate performance; subjective comments from participants showed a preference for the billboard style as well. We therefore suggest, for reading tasks, that designers of interfaces for games, video, and augmented reality provide billboard style to maximize readability for the widest range of applications.

In [4], Bolter and Grusin focus on the relationship between visual digital expressions (such as Virtual Reality) and earlier media forms (written text, photography). They argue that digital forms both borrow from and seek to surpass earlier forms; they give this process the name "remediation". We believe that integrating text with video (e.g., AR applications) and with 3D graphics (e.g., 3D comic books) can be one of the remediations of the electronic text. If we develop new usable ways of presenting such information on screen, which enhance reading activity, if we can provide materials, which exceed current text's limitations, this may ultimately result in better, more effective conveying information, and more engaging education.

Acknowledgments

The work presented in this paper has been funded in part by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 and Enterprise Ireland under Grant No. REI1004.

REFERENCES

1. Azuma, R. and Furmanski, C. Evaluating Label Placement for Augmented Reality View Management, In *Proc. ISMAR 2004*.
2. Bell, B., Feiner, S., and Hollerer, T. View Management for Virtual and Augmented Reality. In *Proc. UIST 2001*.
3. Bernard, M.L., Chaparro, B.S., Mills, M.M. and Halcomb, C.G. Comparing the effects of text size and format on the readability of computer-displayed Times New Roman and Arial text. *IJHCS*, 2003.
4. Bolter, J.D. and Grusin, R. *Remediation: Understanding New Media*, 2000.
5. Chen, J., Pyla P., and Bowman D., Testbed Evaluation of Navigation and Text Display Techniques in an Information-Rich Virtual Environment. In *Proc. Virtual Reality 2004*.
6. Darroch, I., Goodman, J. Brewster, S. and Gray, P. The Effect of Age and Font Size on Reading Text on Handheld Computers. In *Proc. INTERACT 2005*.
7. Digital television recommendations: http://www.dtg.org.uk/ and http://www.ofcom.org.uk/
8. Dillon, A. Reading from paper vs. screens: A critical review of the empirical literature. *Ergonomics*, 1992.
9. Feiner, S., Augmented Reality: A new way of seeing, *Scientific American*, 2002.
10. Gabbard, J.L., Swan, J.E. and Hix, D. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence*, 2006.
11. Gould, J.D., Alfaro, L., Finn, R., Haupt, B. and Minuto, A. Reading from CRT displays can be as fast as reading from paper. *Human Factors*, 1987.
12. Harrison, B.L. and Vicente, K.J. An experimental evaluation of transparent menu usage. In *Proc. CHI 1996*.
13. Hill, A.L. and Scharff, L.F. Readability of computer displays as a function of color, saturation, and background texture. *Engineering Psychology & Cognitive Ergonomics*, 1999.
14. Jankowski, J. and Decker, S. 2LIP: Filling the Gap between the Current and the Three-Dimensional Web, In *Proc. Web3D 2009*.
15. Jorna, G.C. and Snyder, H.L. Image Quality Determines Differences in Reading Performance and Perceived Image Quality with CRT and Hard-Copy Displays. *Human Factors*, 1991.
16. Larson, K., van Dantzich, M., Czerwinski, M. and Robertson, G. Text in 3D: some legibility results. In *Proc. CHI 2002*.
17. Lavie, T. and Tractinsky, N. Assessing dimensions of perceived visual aesthetics of web sites. *IJHCS*, 2004.
18. Leykin, A. and Tuceryan, M. Automatic Determination of Text Readability over Textured Backgrounds for Augmented Reality Systems. In *Proc. ISMAR 2004*.
19. Maass, S. and Dollner, J. Embedded labels for line features in interactive 3D virtual environments, In *Proc. Afrigraph 2007*.
20. Nunberg, G. *The Future of the Book*, 1997
21. Polys, N.F., Kim, S. and Bowman, D.A. Effects of Information Layout, Screen Size, and Field of View on User Performance in Information-Rich Virtual Environments. In *Proc. VRST 2005*.
22. Preim, B., Raab, A., and Strothotte, T. Coherent Zooming of Illustrations with 3D-Graphics and Text. In *Proc. Graphics Interface 1997*.
23. Ritter, F., Sonnet, H., Hartmann, K., and Strothotte, T. Illustrative Shadows: Integrating 3D and 2D Information Displays. In *Proc. IUI 2003*.
24. Riverside Publishing. *Nelson-Denny Reading Test*, 1993.
25. Scharff, L.F., Hill, A.L. and Ahumada, A.J. Predicting the readability of transparent text. *Journal of Vision*, 2002
26. Shneiderman, B. and Plaisant, C. *Designing the user interface: strategies for effective Human-Computer Interaction*, 2004.
27. Sonnet, H., Carpendale, S., and Strothotte, T. Integrating expanding annotations with a 3D explosion probe. In *Proc. AVI 2004*.
28. Sonnet, H., Carpendale, S. and Strothotte, T. Integration of 3D Data and Text: The Effects of Text Positioning, Connectivity, and Visual Hints on Comprehension. In *Proc. INTERACT 2005*.
29. Tractinsky, N., Shoval-Katz, A., and Ikar, D. What is beautiful is usable. *Interacting with Computers*, 2000.